ELSEVIER

Acta Psychologica 114 (2003) 101–114

acta psychologica

www.elsevier.com/locate/actpsy

An own gender bias and the importance of hair in face recognition

Daniel B. Wright [a,*], Benjamin Sladden [b]

[a] *Psychology Department, University of Sussex, Brighton BN1 9QH, UK*
[b] *University of Bristol, UK*

Received 8 November 2002; received in revised form 13 May 2003; accepted 3 June 2003

Abstract

There is a large literature on the *own race bias*, the finding that people are better at recognizing faces of people from their own race. Here an *own gender bias* is shown: Males are better at identifying male faces than female faces and females are better at identifying female faces than male faces. Encoding a person's hair is shown to account for approximately half of the own gender bias when measured using hit and false alarm rates. Remember/know judgements and confidence measures are taken. Encoding a person's hair is critical for having a "remember" recollective experience. Parallels with the own race bias and implications for eyewitness testimony are discussed.

PsycINFO classification: 2343
Keywords: Memory; Face recognition; Gender bias; Own race bias; Eyewitness

1. Introduction

Ninety percent of eyewitness experts believe that the *own race bias*—that people are better at identifying others of their own race—has produced results that are reliable and large enough to be part of expert scientific testimony (Kassin, Tubb, Hosch, & Memon, 2001; see special issue of *Psychology, Public Policy, and Law*, March 2001; for a meta-analysis see Meissner & Brigham, 2001). The explanation usually given for this bias is that people become experts in recognizing faces of their own race because of having much interest in and contact with people of their own

* Corresponding author.
E-mail address: danw@sussex.ac.uk (D.B. Wright).

0001-6918/$ - see front matter
doi:10.1016/S0001-6918(03)00052-0

race (Brigham & Malpass, 1985; Furl, Phillips, & O'Toole, 2002; Wright, Boyd, & Tredoux, 2003). Other own group biases, for example, an own age bias, have also been found (Wright & Stroud, 2002). Here we are interested in an own gender bias, and the quality of the resulting memories.

There have been some studies that have investigated whether there is an own gender bias. For example, Shaw and Skolnick (1994, 1999) have conducted several studies investigating this. In their studies, male and female participants were shown simulated crime videos of either a male or a female perpetrator. They found that people were more accurate with their own gender. This type of study, where participants see a small number of targets in a relatively realistic format, is important for establishing the ecological validity of any finding. However, there are two disadvantages of this type of study which we address here. The first is that each person can contribute only a small number of data points. This means that a large number of participants is needed to achieve precise estimates and to have sufficient power to detect even moderate effects. The second disadvantage is that only a small number of faces are used as stimuli. There may be peculiarities about these particular faces that make the results not generalize (for discussion see Wells & Windschitl, 1999; Wright, 1998). The standard "old"/"new" memory recognition procedure does not have these disadvantages, though it has less ecological validity than the approach used by Shaw and Skolnick (1994, 1999). Thus, these studies should be seen as complementary to Shaw and Skolnick's approach.

The results of studies using "old"/"new" recognition procedures have been mixed (McKelvie, 1981, 1987). Shapiro and Penrod (1986) conducted a meta-analysis of a large number of face recognition studies. Some of these studies reported the gender of the face and the participant, so this allowed the authors to look to see if there was, overall, an effect. Shapiro and Penrod found an own gender bias for correct identifications, but no bias for correctly rejecting faces that had not been previously seen. The effect, however, was much smaller than for the own race effect, and the size/presence of the effect varied across studies. The effect also does not appear to occur to the same extent for females and males. In one early study, Cross, Cross, and Daly (1971) found the effect was due mostly to females performing better with female faces than male faces, not to a similar own gender bias for male participants (see also Lewin & Herlitz, 2002).

As with other own group biases, the own gender bias is shown by a statistical interaction. Here it is between the gender of the target face and the gender of the participant. Herlitz and colleagues (Herlitz, Airaksinen, & Nordstrom, 1999; Lewin, Wolgers, & Herlitz, 2001; Nyberg, Habib, & Herlitz, 2000) have shown gender differences in memory performance on several test batteries. Females are generally better with verbal tasks and males are generally better with visuospatial tasks. With respect to face recognition, these gender biases would show up as main effects (Lewin & Herlitz, 2002). It is more difficult to explain an own gender interaction from gender differences on these test batteries unless characteristics of female faces differ in their ease of verbal encoding compared with male faces.

A second aim of this research is to investigate memory for a person's hair. O'Donnell and Bruce (2001) have shown that hair is a particularly important characteristic for recognizing unfamiliar faces. This is of great applied importance because criminals often cover their hair or use other disguises when committing crimes. By simultaneously examining own gender bias and memory for hair we are able to estimate the proportion of any observed own gender bias that is attributable to memory for hair. It might be, for example, that males and females are equally accurate trying to recognize someone who had his or her hair covered at encoding. This would mean that the entire own gender bias could be attributed to a bias in memory for hair. Alternatively it may be that the bias remains the same size when hair is covered, suggesting hair does not contribute to the own gender bias.

A third aim of this study involves people's recollective experience of the memory. We use an "old"/"new" memory recognition procedure with Tulving's (1985; see also Gardiner, 1988) "remember"/"know" distinction. Tulving (1985) said that these responses represent different levels of memory awareness. "Remember" responses involve what he calls autonoetic, or self-awareness, and equate with episodic memory. "Know" responses are where the person has a memory of the event, but without the "I was there" recollective experience which is associated with "remember" responses. This is similar to semantic memory and involves noetic consciousness. Tulving describes a third type of memory which is without consciousness, or anoetic, and corresponds to "guess" responses (see Gardiner, Ramponi, & Richardson-Klavehn (2002), for meta-analysis of studies using these three options). This is often called procedural or implicit memory. Following Conway, Gardiner, Perfect, Anderson, and Cohen (1997), we use an additional category, "familiar," which lies between "know" and "guess." It maps onto some notion of accessibility of aspects of the event, but without any conscious memory per se. The "remember" responses usually coincide with perceptual memories (see participants' instructions below). If hair is necessary to define some holistic face representation, then it is possible that encoding without hair may reduce the number of "remember" judgements. We also asked for decision confidence. Research (Gardiner & Java, 1990) has shown that remember/know judgements differ from confidence ratings, though the two are statistically related (Yonelinas, 2001).

2. Methods

2.1. Participants

Twenty female and 20 male Caucasian students from the University of Bristol volunteered for this study. The ages ranged from 17 to 36 with a mean of 26 years. All were unpaid volunteers.

2.2. Materials

Twenty four male and 24 female face pictures were downloaded from the Web or taken from Bristol psychology department from before any of the participants would have arrived. All were Caucasian. None had any outstanding features (glasses, beards, excessive jewellery, odd hair styles, etc.). Six judges estimated the ages of these faces and the estimates for all faces were between 18 and 25 years. In general, people are accurate in estimating people's ages (George & Hole, 1995). Using a computer graphics package, these pictures were altered to remove any additional non-facial cues. Each picture was printed twice and mounted on a card. The hair was removed using a black permanent ink marker from one of the pictures for each person.

2.3. Procedure

Participants were tested individually and told that they would be seeing 24 faces and that half of the faces would be shown with their hair blacked out. Twenty four faces were randomly chosen with the constraint that there were six males with hair, six males with no hair, six females with hair, and six females with no hair. Participants were shown each of the target faces, in random order, at a rate of 4 s per face. The cards were lifted up by the experimenter (B.S.) who could not see the front of the card.

The instructions for the recognition phase, recollective state judgements, and confidence ratings were then given to the participant. The recognition was simply saying "yes" if they thought that they had seen the face before and "no" if they had not. Again, the experimenter lifted the card and recorded responses, but could not see the front of the card. Recollective states were defined as follows:

Remember — You can recall the specific episode of first viewing the face. In this case you may have the image in mind relating to the recalled face. Perhaps you virtually "see" again the original presentation of the face.

Know — You may just feel that you recognize the face from the presentation stage. In this case you would not be able to recall a specific episode, you simply know the face.

Familiar — It may be that you neither know the face nor remember it, although the face seems more familiar than other faces which you previously responded "no" to.

Guess — You will have seen 50% of the faces previously in the presentation stage. From this you may deduce that you have not made the response "yes" enough times and for this reason you select the face. This is a guess response and is usually accompanied by a lack of confidence.

Confidence ratings were made on a 1–10 scale with 1 being a lack of confidence and 10 being certain.

Participants were told to ask the experimenter if they had any questions. Explaining this procedure took approximately 5 min. Thus, the retention interval from when the last face was shown until the beginning of testing was 5 min. Next, participants were presented with the 24 male and 24 female pictures, all with hair, in random order. They were first asked whether they had seen the face, then the confidence in their decision, and finally, if they had said "yes," their recollective state. Afterwards, participants were thanked and debriefed.

3. Results

3.1. Recognition scores

Fig. 1 shows the hit rate for male and female faces originally shown with and without hair, and the false alarm rate for new male and new female faces. The hit rates and false alarm rates are the number of "yes" responses plus 0.5 divided by the number of chances plus one. The flattening constant was used so that z scores can be calculated when the hit or false alarm rate is either 0 or 1. There is only a single set of false alarm rates because the photographs not previously shown, and therefore eligible to be false alarms, cannot vary in the way they were previously shown. In all comparisons, people performed better with their own gender demonstrating the own gender bias. The size of the effects are largest for the hit rate when the faces were originally shown with hair.

For statistical analyses the hit and false alarm rates were combined into d' scores (z(hit rate)-z(false alarm rate), see Macmillan, 1993). Using the confidence ratings, we calculate receiver operating characteristics (ROCs) for the different conditions using the procedures described in Wickens (2002, Chapter 5). When graphed as their normal deviates all were roughly linear with slopes of 1. These are assumptions of d'. We choose d' rather than alternatives (Swets, 1986) because of these ROCs and because it is the most well known measure for diagnostic accuracy.

An ANOVA was run on the d' scores (Table 1) with two within-subject variables (target gender and whether the face was originally shown with hair) and one

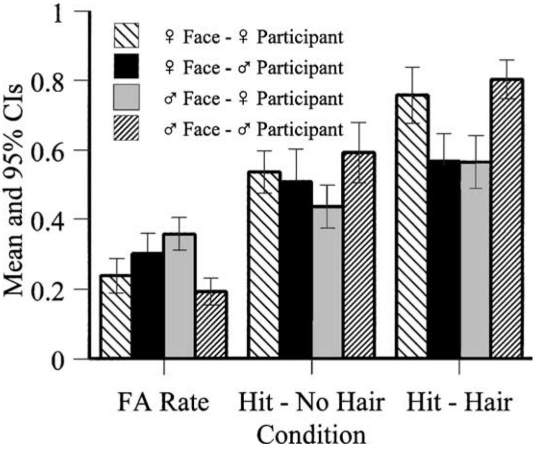

Fig. 1. The means and 95% confidence intervals (CIs) for the hit and false alarm rates for the different conditions.

Table 1
The d' scores for male and female faces, for male and female participants, when hair is present at encoding and when it is covered

d'				
	Male face		Female face	
	No hair	Hair	No hair	Hair
Males	1.18 (0.15)	1.86 (0.13)	0.56 (0.21)	0.76 (0.16)
Females	0.21 (0.10)	0.56 (0.13)	0.85 (0.10)	1.56 (0.17)

Scores of $d' = 0$ correspond to no accuracy. Standard errors are in parentheses below the estimates.

between-subject variable (participant's gender). The most notable effects are that the d's were higher when hair was present ($F(1,38) = 49.61$, $p < 0.001$, partial $\eta^2 = 0.57$) and that male and female participants each did better with their own gender ($F(1,38) = 72.98$, $p < 0.001$, partial $\eta^2 = 0.66$), thus demonstrating the own gender bias. The magnitude of this effect is approximately the same for males (main effect of target gender for males, partial $\eta^2 = 0.63$) and females (partial $\eta^2 = 0.70$). Finally, there was a three-way interaction among participant's gender, target's gender, and whether the target's hair was covered ($F(1,38) = 14.39$, $p = 0.001$, partial $\eta^2 = 0.28$). The reason for this is that each group had a larger "hair" effect for their own gender. In other words, not encoding hair is more detrimental for faces of your own gender than for faces of the other gender.

Pairwise comparisons between d' for faces shown with and without hair were conducted for male and female participants, for both male and female photographs. The 95% between-subject confidence intervals for the differences (as opposed to the d's themselves), which provide measures of the benefit of originally encoding hair, are:

males viewing female faces (0.07, 0.46),
males viewing male faces (0.41, 0.94),
females viewing female faces (0.44, 0.97),
females viewing male faces (0.11, 0.60).

As can be seen, the mean hair effect is about twice as large when viewing people from your own gender. It is important to note that all the confidence intervals are above zero, meaning that the hair effect is detected in each condition, it is just smaller when recognizing people of the other gender.

There is an alternative interpretation of the interaction in Table 1. The ANOVA statistics are based on additive differences between conditions: d' with hair minus d' without hair. This assumes a particular level of measurement (interval). When analysed this way the hair shift in d' is about 0.70 for males and females when viewing faces of their own gender, and is about 0.20–0.30 when viewing faces of the other gender. However, it is important to consider other levels of measurement (Wright, 1997a). If instead of taking the additive difference, the ratio is taken, then the hair effects for male participants are $1.86/1.18 = 1.58$ for male faces and $0.76/0.56 = 1.36$ for female faces. The effects for female participants are $0.56/0.21 = 2.67$ for males faces and $1.56/0.85 = 1.84$ for female faces. This alternative interpretation suggests that female participants were more affected by the absence of hair than male participants and that male faces are more affected by having hair covered than female faces. Here we focus on the additive difference interpretation because it is more common when using d' in recognition studies, but this deserves further investigation.

3.2. Quality of responses and confidence

Table 2 gives the frequency and percentage (in parentheses) for the faces the person says are new, and those which the person responds "remember," "know," "familiar," and "guess," for the different conditions. These are broken down by whether the recognition was of someone of their own gender or of someone from the other gender. In total, 0.2% (two in total) of those faces not originally shown (i.e., fillers), 1.5% (seven in total) of those originally shown without hair, and 22.4% (108 in total) of those originally shown with hair were "remembered." This very large difference between those shown with and those shown without hair was maintained across

Table 2
Participants' reported recollective state by condition for cross gender and own gender recognition

	Recollective state					
	Says new	Remember	Know	Familiar	Guess	Total
Cross gender recognition						
Not seen	328	1	27	102	21	479
	(69)	(0.2)	(6)	(21)	(4)	
Seen w/o	128	4	27	57	24	240
Hair	(53)	(2)	(11)	(24)	(10)	
Seen w/	102	39	53	40	7	241
Hair	(42)	(16)	(22)	(17)	(3)	
Totals	558	44	107	199	52	960
	(58)	(5)	(11)	(21)	(5)	
Own gender recognition						
Not seen	388	1	17	60	14	480
	(81)	(0.2)	(4)	(13)	(3)	
Seen w/o	101	3	32	95	7	238
Hair	(42)	(1)	(13)	(40)	(3)	
Seen w/	42	69	70	55	6	242
Hair	(17)	(29)	(29)	(23)	(3)	
Totals	531	73	119	210	27	960
	(55)	(8)	(12)	(22)	(3)	

The frequencies are shown with the row percentages in parentheses. The first row gives the figures for previously unseen faces, the second row gives faces previously seen without hair, and the third row those previously seen with hair.

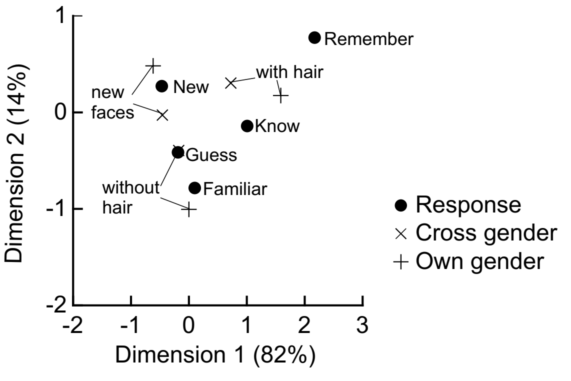

Fig. 2. A biplot showing the associations among own versus other gender recognitions, originally encoding a face either with or without hair, and recollective experience. This is based on a correspondence analysis (see Appendix A for further details).

genders. The "know" responses showed a similar trend. The "familiar" responses showed the opposite trend. The "guess" option was seldom used (4.2% in total).

Clearly the memory quality ratings are related to the condition. Assuming that the individual trials are independent, the association between condition and response can be measured by $\chi^2(8) = 155.7$ for recognition of people of the other gender and $\chi^2(8) = 451.8$ for recognition of people of the same gender. This association can be partitioned in several ways. Fig. 2 shows the biplot from a correspondence analysis. Correspondence analysis is a method for partitioning the association of two variables and producing a graph of the associations. Items similar to each other will tend to be near to each other in the graph. Details of the statistical procedure are given in Appendix A.

First looking at the responses, Fig. 2 shows a progression from the guess and familiar responses through know responses to remember responses. This progression shows that recognition of faces seen previously without hair are associated with guess and familiar responses, while recognitions of those previously seen with hair are associated with know and remember responses. Faces not previously seen (new faces) are associated with new responses. All the cross gender conditions (shown with the symbol ×) are all closer to the origin (the point 0, 0) than the own gender conditions (shown with the symbol +). This means that the own gender conditions are differentiated better by the response conditions. In essence, this means that people are more accurate with own gender trials.

Confidence ratings were also made. Excluding those who said that the face was not previously seen, and treating all observations as independent, the γ correlation was 0.87 ($p < 0.001$). The pattern was similar for all conditions when making a positive recognition. However, when people said that the face was not previously seen, there were differences among the conditions. When it was a correct rejection people were fairly confident ($\bar{x} = 6.16$, sd = 1.87), while they were less confident when incorrectly rejecting a face originally shown without hair ($\bar{x} = 5.21$, sd = 1.74), and least confident when rejecting a face originally shown with hair ($\bar{x} = 4.69$, $\text{sd} = 1.81$). Assuming item independence, these differences are all highly significant.

4. Discussion

The aims of this study were to test if an own gender bias exists, to see if the size of any own gender bias is related to encoding hair features, and to examine how having hair covered at encoding affects recollective experience. All these were found: an own gender bias, hair contributed to this effect, and a very large effect of hair on recollective experience.

A strong own gender bias was found. This is important for several reasons. One is related to eyewitness testimony. Police investigators and jurors should be warned of this effect. While the own race bias has an intuitive explanation, that it is caused by contact time and levels of processing of one's own race, this seems less intuitive an explanation for an own gender bias. However, from an evolutionary standpoint it is likely that our ancestors were more concerned with recognizing their competition for mating than recognizing possible mates (Karen McComb, personal communication, January, 2003). Further, it is clear from non-pornographic magazines, that the majority of photographs are of people of the same gender as the target audience. While exposure may still be a partial explanation for the own gender bias, it deserves further investigation.

The size of the effect is also important. In fact, according to the rules for admitting expert scientific testimony as evidence (Daubert v Merrell Dow Pharmaceutical, Inc., 1993; and Kumho Tire Co., v. Carmichael, 1999), the judge is to consider the size of any effects. With respect to the quantitative measure d', the own gender bias was of a similar size to not encoding a person's hair. Unlike Shapiro and Penrod (1986), we found evidence for a recognition bias for false alarms. Further, in contrast with some of the literature (Lewin & Herlitz, 2002), we found the bias occurred for both male and female faces.

The second aim of this study was to see if encoding information about a person's hair could have differential benefits for male and female faces, and by male and female participants. We did not have any a priori hypotheses about whether hair would be a more diagnostic feature of male or female faces, or if males or females would differentially use characteristics of the targets' hair to help in memory. These effects could be examined in several ways. The most common method is subtracting the d' for without hair recognition from with hair recognition, and comparing these differences. Using this method, we found a three-way interaction with gender of face, gender of participant, and presence of hair. The direction of this effect was that hair was more helpful for females viewing female faces, and males viewing male faces, than when making cross gender identifications. Thus, memory of hair appears to be a contributing factor to the own gender bias. From the effect size measures, about half of the own gender bias can be accounted for by memory for hair. This method assumes that the d' scores have an interval level of measurement. This is the usual assumption, but other metrics are possible (Wright, 1997a). If a ratio of the d' scores was thought to be the most meaningful level of measurement the interpretation would be different. Further investigations into which the appropriate level of measurement and also whether other facial features contribute to this effect are warranted.

The third focus of this study was on the recollective memory experience (Tulving, 1985). While the quantitative measure d' showed that encoding hair was important, the qualitative measures showed more striking differences. People were not "remembering" faces which were previously seen without hair. While participants said they "remembered" the face between 20% and 25% of the time when originally shown with hair, only approximately 2% of the time did they say they "remembered" the face when they had not seen the person's hair. This is a very large effect. The applied significance of this is that police, judges, and jurors should not expect a vivid recollection when the culprit's hair is covered during the crime.

In summary, research into biases in face recognition is important both for understanding the psychological mechanisms involved and for applied purposes. Here we have shown that an own gender bias occurs in face recognition. This is important for assessing the reliability of eyewitness reports. Some changes to cross gender lineup procedures could be made along the lines of those advocated for cross race lineup procedures (Wells & Olson, 2001). These could include making sure that there are people of the same gender there to advise police on the construction of the lineup and that more fillers are used for cross gender lineups. We showed that the own gender bias is due partially to memory for the person's hair. This also has forensic relevance because often people cover their hair when committing a crime. They do this because they believe that it makes them less easy to recognize. We confirmed this. The size of this effect, when looking just at the quantitative measures (hit rates, false alarm rates, and d'), is of a similar size to the own gender bias. However, the effect is much more striking when looking at qualitative measures of the recollective experience. These were greatly affected by removing hair from the encoded faces.

Appendix A

Correspondence analysis is a relatively uncommon technique in psychology and so it deserves further explanation here. We also go into more depth about some of the models that we explored for comparing memory quality and the different conditions.

Correspondence analysis is appropriate when there are two categorical variables and the analyst is interested in the association between them. Here, there are four variables, but correspondence analysis can be used by combining some of the variables (Van der Heijden, de Falguerolles, & de Leeuw, 1989; Wright, 1997b, Chapter 7). Other techniques can also be used with multiple categorical variables, but correspondence analysis provided a better description of the data here. Introductions to correspondence analysis can be found in Bartholomew, Steele, Moustaki, and Galbraith (2002) and Michailidis and De Leeuw (1998). More detailed treatment is given in Van de Geer (1993). It is also worth mentioning the main alternative to correspondence analysis in this situation, log-multiplicative models (the RC(M*) models, see Agresti, 2002; Goodman, 1991). The two approaches yield, as they usually do, similar solutions for the current data. Our choice of correspondence analysis was because we wished to stress the descriptive nature of our analyses.

First, it is important to show how the number of variables was reduced from 4 to 2 so that correspondence analysis could easily be used. There are four variables of interest in this data set: the response (five categories: remember, know, familiar, guess, and new), the gender of the person (two categories), the gender of the face (two categories), and trial condition (three categories: not previously shown, previously shown with hair, previously shown without hair). As reported in the text, the effects for males and females (participants and faces) are of approximately the same size. Thus, these two variables were combined into one variable: own gender versus cross gender trials. Next, this own gender variables was combined with the condition into six categories for each possible combination. This corresponds to the six rows in Table 2.

The most common procedure for comparing two categorical variables is a χ^2 test of association. For the data in Table 2 this is $\chi^2(20)$ of 666.25 for Pearson χ^2. There are two main questions: How big is this effect? and what is the nature of any association? The first of these can be addressed with the χ^2 value. The maximum Pearson χ^2 is n multiplied by the number categories, minus 1, of the variable with the fewer number of categories. In symbols, $n \min(r-1, c-1)$. So, here the maximum is $(1920)(4) = 7680$. An intuitive effect size is the proportion of observed χ^2 of the possible χ^2. This is 0.087 and is sometimes called V^2. Its square root, Cramer's $V = 0.29$, is another common measure of effect size.

The χ^2 value shows that the independence model of no association does not fit well (i.e., it is statistically significant). But with 20 degrees of freedom the statistic does not identify where the association lies and so, on its own, is of little value. There are different ways to explore the association, for example, graphing the residuals from the independence model. However, in recent years there have been two strands of research that appear more useful. The first set, which includes correspondence analysis, is built on ideas from scaling and principal component analysis. The second set extends the log-linear model, and these techniques are more popular in sociology. Goodman (1991) and others have described the similarities and differences between these approaches.

Correspondence analysis is a descriptive method of partitioning the Pearson χ^2 association into a small number of dimensions, or canonical correlates. The qualitative values for the variables are assigned quantitative values. The quantitative values found for each value maximize the correlation between the two variables. This is the first dimension. Here the first canonical correlate is created by assigning the following values to the response variable (new = −0.47, guess = −0.19, familiar = 0.21, know = 1.01, remember = 2.17) and these values to the conditions (cross gender new = −0.46, own gender new = −0.62, cross no hair = −0.18, own no hair = 0.00, cross with hair = 0.72, own with hair = 1.59). If these values were assigned to the two variables, the correlation would be $r = 0.54$.

Further dimensions can be calculated up to the minimum of the number categories of each variable, minus 1 (i.e., $\min(r-1, c-1)$). The values for the second dimension are: for the response variable (new = 0.57, guess = 0.04, familiar = 0.21, know = 0.12, remember = 0.06) and for the conditions (cross gender new = −0.03, own gender new = 0.48, cross no hair = −0.40, own no hair = −1.01, cross hair = 0.30, own hair = 0.18). These produce a correlation of 0.22. The scores on the first two dimensions can be plotted together as a biplot as in Fig. 2. The third dimension produces $r = 0.11$, and the fourth $r = 0.04$. Squaring the correlation yields what is called in correspondence analysis the inertia for each dimension and these values can be used to decide the number of dimensions in the same manner as a scree plot of the eigenvalues used in factor analysis to decide the number of factors (the sum of the inertia of all the dimensions is χ^2/n). However, usually two dimensions are plotted.

The third dimension differentiates guesses from the other responses and the cross gender no hair condition from the other conditions. The two dimensional solution underestimates the association between guesses and this condition. The fourth dimension accounts for too little of the total inertia to be interpreted.

It appears that the guess responses may involve more than just a different point along a memory quality dimension. This supports the conclusions from Gardiner et al.'s (2002) meta-analysis. Correspondence analysis was rerun without the trials where the participant said that an item was "new." The new Pearson $\chi^2(15) = 245.51$ with $V^2 = 0.098$, $V = 0.32$. The first dimension produced a correlation of $r = 0.52$, the second an $r = 0.16$, and the third an $r = 0.05$. The biplot (Fig. 3) shows the first two dimensions. Remember, know, and familiar form an almost interval scale (in relation to the responses to the different conditions). The second dimension differentiates the guess responses from the others, as well as showing a differentiation between cross and own gender faces shown without hair. The own gender trials were more associated with familiar responses, but the cross gender trials were more asso-

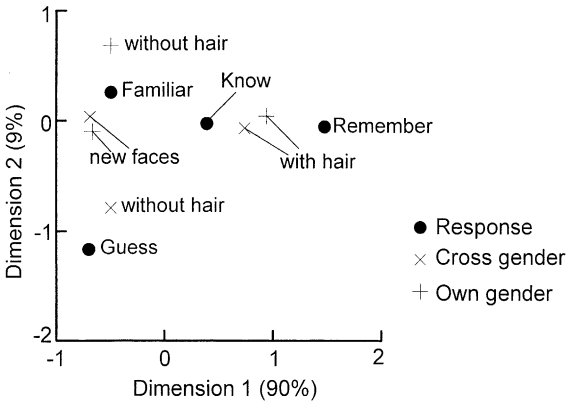

Fig. 3. A biplot of the associations among own versus other gender recognitions, originally encoding a face either with or without hair, and recollective experience, with the "new" responses excluded.

ciated with guess responses. This is clearly an exploratory observation, but deserves further investigation. In general, we encourage others to use techniques for multivariate categorical data with memory quality responses in order to help understand how these responses relate to memory. Given the amount of research using categorical responses for memory quality (i.e., remember–know judgements, source monitoring, etc.), fuller exploration of the properties of these scales would be prudent. For these data we found correspondence analysis the most informative.

References

Agresti, A. (2002). *Categorical data analysis* (2nd ed.). Hoboken, NJ: John Wiley & Sons, Inc.

Bartholomew, D. J., Steele, F., Moustaki, I., & Galbraith, J. I. (2002). *The analysis and interpretation of multivariate data for social scientists*. Boca Raton, FL: Chapman & Hall/CRC.

Brigham, J. C., & Malpass, R. S. (1985). The role of experience and contact in the recognition of faces of own-race and other-race persons. *Journal of Social Issues, 41*, 139–155.

Conway, M. A., Gardiner, J. M., Perfect, T. J., Anderson, S. J., & Cohen, G. M. (1997). Changes in memory awareness during learning: the acquisition of knowledge by psychology undergraduates. *Journal of Experimental Psychology: General, 126*, 393–413.

Cross, J. F., Cross, J., & Daly, J. (1971). Sex, race, age, and beauty as factors in recognition of faces. *Perception & Psychophysics, 10*, 393–396.

Daubert v Merrell Dow Pharmaceutical, Inc. (1993). 113 S CT. 2786.

Furl, N., Phillips, P. J., & O'Toole, A. J. (2002). Face recognition algorithms and the other-race effect: computational mechanisms for a developmental contact hypothesis. *Cognitive Science, 26*, 797–815.

Gardiner, J. M. (1988). Functional aspects of recollective experience. *Memory & Cognition, 16*, 309–313.

Gardiner, J. M., & Java, R. I. (1990). Recollective experience in word and nonword recognition. *Memory & Cognition, 18*, 23–30.

Gardiner, J. M., Ramponi, C., & Richardson-Klavehn, A. (2002). Recognition memory and decision processes: a meta-analysis of remember, know, and guess responses. *Memory, 10*, 83–98.

George, P. A., & Hole, G. J. (1995). Factors influencing the accuracy of age-estimates of unfamiliar faces. *Perception, 24*, 1059–1073.

Goodman, L. A. (1991). Models, measures, and graphical displays in the analysis of contingency tables (with discussion). *Journal of the American Statistical Association, 86*, 1085–1138.

Herlitz, A., Airaksinen, E., & Nordstrom, E. (1999). Sex differences in episodic memory: the impact of verbal and visuospatial ability. *Neuropsychology, 13*, 590–597.

Kassin, S. M., Tubb, V. A., Hosch, H. M., & Memon, A. (2001). On the "general acceptance" of eyewitness testimony research: a new survey of the experts. *American Psychologist, 56*, 405–416.

Kumho Tire Co., v. Carmichael (1999). US LEXIS 2189, No. 97-1709, decided 23 March 1999.

Lewin, C., & Herlitz, A. (2002). Sex differences in face recognition: women's faces make a difference. *Brain and Cognition, 50*, 121–128.

Lewin, C., Wolgers, G., & Herlitz, A. (2001). Sex differences favoring women in verbal but not in visuospatial episodic memory. *Neuropsychology, 15*, 165–173.

Macmillan, N. A. (1993). Signal detection theory as a data analysis method and psychological decision model. In G. Keren, & C. Lewis (Eds.), *A handbook for data analysis in the behavioral sciences: methodological issues* (pp. 21–57). Hillsdale, NJ: Lawrence Erlbaum Associates.

McKelvie, S. J. (1981). Sex differences in memory for faces. *Journal of Psychology, 107*, 109–125.

McKelvie, S. J. (1987). Sex differences, lateral reversal, and pose as factors in recognition memory for photographs of faces. *Journal of General Psychology, 114*, 13–37.

Meissner, C. A., & Brigham, J. C. (2001). Thirty years of investigating the own-race bias in memory for faces: a meta-analytic review. *Psychology, Public Policy and Law, 7*, 3–35.

Michailidis, G., & De Leeuw, J. (1998). The Gifi system of descriptive multivariate analysis. *Statistical Science, 13*, 307–336.

Nyberg, L., Habib, R., & Herlitz, A. (2000). Brain activation during episodic memory retrieval: sex differences. *Acta Psychologica, 105*, 181–194.

O'Donnell, C., & Bruce, V. (2001). Familiarisation with faces selectively enhances sensitivity to changes made to the eyes. *Perception, 30*, 755–764.

Shapiro, P. N., & Penrod, S. (1986). Meta-analysis of facial identification studies. *Psychological Bulletin, 100*, 139–156.

Shaw, J. I., & Skolnick, P. (1994). Sex-differences, weapon focus, and eyewitness reliability. *Journal of Social Psychology, 134*, 413–420.

Shaw, J. I., & Skolnick, P. (1999). Weapon focus and gender differences in eyewitness accuracy: arousal versus salience. *Journal of Applied Social Psychology, 29*, 2328–2341.

Swets, J. A. (1986). Indexes of discrimination or diagnostic-accuracy: their ROCs and implied models. *Psychological Bulletin, 99*, 100–117.

Tulving, E. (1985). How many memory systems are there? *American Psychologist, 40*, 385–398.

Van de Geer, J. P. (1993). *Multivariate analysis of categorical data: Applications*. London: Sage Publications.

Van der Heijden, P. G. M., de Falguerolles, A., & de Leeuw, J. (1989). A combined approach to contingency table analysis using correspondence analysis and log-linear analysis [with discussion]. *Applied Statistics, 2*, 249–292.

Wells, G. L., & Olson, E. A. (2001). The other-race effect in eyewitness identification. What do we do about it? *Psychology, Public Policy and Law, 7*, 230–246.

Wells, G. L., & Windschitl, P. D. (1999). Stimulus sampling in social psychological experimentation. *Personality and Social Psychology Bulletin, 25*, 1115–1125.

Wickens, T. D. (2002). *Elementary signal detection theory*. New York: Oxford University Press.

Wright, D. B. (1997a). Football standings and measurement levels. *The Statistician: Journal of the Royal Statistical Society Series D, 46*, 105–110.

Wright, D. B. (1997b). *Understanding statistics: An introduction for the social sciences*. London: Sage Publications.

Wright, D. B. (1998). People, materials, and situations. In J. A. Nunn (Ed.), *Laboratory Psychology* (pp. 97–116). Hove: Lawrence Erlbaum Associates.

Wright, D. B., Boyd, C. E., & Tredoux, C. G. (2003). Inter-racial contact and the own race bias for face recognition in South Africa and England. *Applied Cognitive Psychology, 17*, 365–373.

Wright, D. B., & Stroud, J. N. (2002). Age differences in lineup identification accuracy: people are better with their own age. *Law and Human Behavior, 26*, 641–654.

Yonelinas, A. P. (2001). Components of episodic memory: the contribution of recollection and familiarity. *Philosophical transactions of the Royal Society of London: Series B—Biological Sciences, 356*, 1363–1374.